AD

Technical Memorandum 16-91

Procedural Text: Predictions of Importance Ratings and Recall by Models of Text Comprehension

Carol Bergfeld Mills
Goucher College and American Institutes for Research

Virginia A. Diehl
Western Illinois University

Deborah P. Birkmire
U.S. Army Human Engineering Laboratory

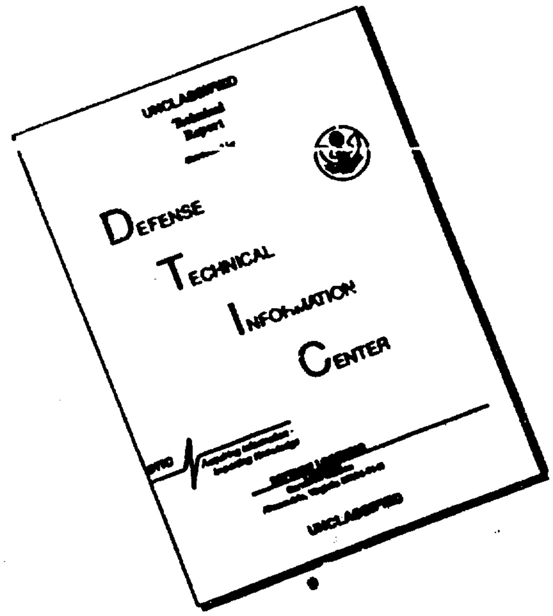
Lien-chong Mou
University of Maryland

August 1991
AMCMS Code 611102.74A0011

Approved for public release;
distribution is unlimited.

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Memorandum 16-91	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Human Engineering Laboratory 6b. OFFICE SYMBOL (If applicable) SLCHE	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21005-5001	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
61102A	1L161102B74A		

11. TITLE (Include Security Classification)
Procedural Test: Predictions of Importance Ratings and Recall by Models of Text Processing

12. PERSONAL AUTHOR(S)
Mills, Carol B.; Diehl, Virginia A.; Birkmire, Deborah P.; Mou, Lien-chong

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ________ TO ________	1991 August	47

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	text processing
05	08		procedural text
			text memory and ratings

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Two models of text comprehension, a process model proposed by Kintsch and van Dijk (1978) and a causal model proposed by Trabasso and Sperry (1985), were tested in two experiments with eight procedural texts. In Experiment 1, 24 female college students rated the importance of propositions, idea units, and sentences to the overall procedure described in the texts on a 7-point scale. In Experiment 2, 16 female college students recalled each of the eight texts immediately after reading it. Predictors derived from the models were used to predict the ratings and the recall in multiple regression analyses. The results showed that the amount of variance accounted for by the predictors varied from text to text. In general, for the ratings and recall, the causal model accounted for more variance than the process model did. For the process model, level within the hierarchy accounted for the most variance and for the causal model, the causal chain factor accounted for the most variance. The models were not as powerful as in previous research with narrative texts perhaps because these models do not take into account the cognitive representation involved in doing the task described by a procedural text.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Technical Reports Office	22b. TELEPHONE (Include Area Code) (301) 278-4478 22c. OFFICE SYMBOL SLCHE-SS-TSB

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

AMCMS Code 61110274A0011 Technical Memorandum 16-91

Procedural Text: Predictions of Importance Ratings and Recall by Models of Text Comprehension

Carol Bergfeld Mills
Goucher College and American Institutes for Research

Virginia A. Diehl
Western Illinois University

Deborah P. Birkmire
U.S. Army Human Engineering Laboratory

Lien-chong Mou
University of Maryland

August 1991

APPROVED: JOHN D. WEISZ
Director
Human Engineering Laboratory

Approved for public release;
distribution is unlimited.

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005-5001

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of Lori Driscoll and Vashti McCollum in data collection in Experiment 1 and of Kathy Vestivich in Experiment 2. Vashti McCollum also calculated the reliabilities for the rating data. The authors would also like to thank Cara Forrest, Sharon Mentro, Susan Rogers, and Tracey Toskes who scored the recall data in Experiment 2.

CONTENTS

PROCEDURAL TEXT: PREDICTIONS OF IMPORTANCE RATINGS AND RECALL BY MODELS OF READING COMPREHENSION

INTRODUCTION

Research has demonstrated that some parts of text materials are more important to readers and are more easily recalled than others. Text units that are rated more important are recalled better than text rated as less important for narratives (Brown & Smiley, 1977; Johnson, 1970). Parts of text placed higher in hierarchical outlines are recalled better, are rated as more important, and are summarized more frequently than those placed lower in the outline (Meyer & McConkie, 1973). These findings have stimulated the development of a number of models of text comprehension, whose goals are to predict which parts of text are more important and are recalled better.

The present research tests two of these models of text comprehension with procedural text, which instructs how to perform an action (e.g., how to make a toy, how to use an electronic device, or how to complete a form), rather than with narratives, which are usually used in this type of research. To test the models, predictor variables were derived for each of the two models. These predictor variables were then used to predict importance ratings and recall provided by readers. This research addresses the issue of the generalizability of the models to a different genre of text. Using the same models on a different genre of text (procedural text) allows comparisons with previous research (Meyer & Rice, 1984). It is well known that there are many differences between genres of text (Graesser & Goodman, 1985), but how these differences affect the fit of the models is not clear from previous research. Testing these models of text comprehension with procedural text will advance one's knowledge about the generality of the models, as well as one's understanding of procedural text, a genre of text that has been studied very little.

The two models tested in this research are the model proposed by Kintsch and his colleagues (Kintsch, 1985; Kintsch & van Dijk, 1978; Miller & Kintsch, 1980; van Dijk & Kintsch, 1983), which will be called the process model, and the model proposed by Trabasso and his colleagues (Trabasso, Secco, & van den Broek, 1984; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985), which will be called the causal model. These models, which have stimulated much recent research, represent two very different theoretical approaches to how text is comprehended and represented.

According to the model developed by Kintsch and his colleagues, differences in reading comprehension of textual information result from processes that occur because of the limited capacity of short-term memory. According to the model, text is parsed into a propositional representation. Then, because of the limitations of short-term memory, the propositions are processed in a series of cycles. Generally, the propositions that are in a sentence are processed together in a cycle. During a cycle, propositions are connected to those already in short-term memory on the basis of argument overlap. If a proposition cannot be connected to those in short-term memory or to those already connected in that specific cycle, a proposition from a previous cycle must be reinstated. If there is no argument overlap with any previous proposition in the text, an inference must be made. When either a reinstatement or an inference must be made, it results in processing difficulties in the cycle.

At the end of the cycle, some of the propositions are carried over in short-term memory to connect to the propositions in the next cycle, allowing the reader to maintain the perception of text coherence. The more times a proposition is carried over in short-term memory, the more likely it will be recalled later. Kintsch and his associates proposed strategies for selecting the information that is carried over from one processing cycle to the next, so that higher level "important" propositions would be retained longer in short-term memory. In addition to this microprocessing component of the model, which has been described, there is also a macroprocessing component, which involves the use of world knowledge to organize the text elements into global concepts. While macroprocessing undoubtedly influences reading, the present research concentrates on microprocessing in procedural text. Previous research (Kintsch & van Dijk, 1878; Miller & Kintsch, 1980) has shown that microprocessing alone is predictive of reading performance.

Research support for the process model comes from a number of studies. Kintsch and Keenan (1973) found that the hierarchical relationships among propositions in isolated sentences as specified in a processing cycle were a powerful determinant of recall, with the higher level propositions being recalled better. For an informally written research report, Kintsch and van Dijk (1978) found a good fit between the predictions from the model and recall and summarization data. Miller and Kintsch (1980) found that for paragraphs taken from a popular magazine, predictions from the microstructure processing of text propositions predicted readability (reading time per proposition recalled), reading time, and recall with the number of inferences and reinstatements being particularly important factors.

According to the causal model, the second model tested in the present research (Trabasso, Secco, & van den Broek, 1984; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985), the importance of a text unit is determined by two factors. The first factor is the number of causal relationships that a text unit has with others in the text. The more causal relationships a given unit has with other units, the more important it is and the better it will be recalled. Causal relationships are determined by using the logical criteria of necessity and counterfactual reasoning, which are described later in this report. The second factor is whether the unit falls on the causal chain. If a unit of text lies on a continuous chain of causal events that connects the opening event and the final outcome, it will be more important and better recalled than if it does not lie on this chain.

Research has shown that both the number of connections and the membership on the causal chain affect performance. However, the relative contribution of these two factors has differed in different experiments. Trabasso, Secco, and van den Broek (1984) found that for the combined data of four stories, units on the causal chain were recalled better than those not on the causal chain, while the number of causal connections had only a minor effect on recall. For six folk tales, Trabasso and Sperry (1985) found judged importance of statements increased with both causal chain membership and number of causal connections, with the number of causal connections uniquely accounting for a significant portion of variance. For five stories, Trabasso and van den Broek (1985) found that membership on the causal chain and the number of causal connections accounted for a significant proportion of variance for four different measures: immediate recall, delayed recall, summarization, and judged importance of the unit. Overall, the causal chain accounted for the most variance in all of the measures. Van den Broek and Trabasso (1986) showed that having membership on the causal chain or having more causal relations with other statements increased the likelihood that an event would be in a summary of a story. Van den Broek (1988), using stories, found that as the number of causal relationships increased, importance ratings increased.

Research provides support for both the process and the causal model; however, the models are obviously different from each other in several ways. First, different mechanisms are held to be responsible for the differences in comprehensibility, recallability, and importance. For the process model, it is assumed that time in short-term memory is critical whereas for the causal model, the causal structure of the text is critical. Second, the process model assumes that only those things that coexist in the limited capacity short-term memory can be connected to each other (without extra processing), whereas with the causal model, any part of a text may be causally related to any other part of the text. Finally, the analyses are based on very different size units of text, with the process model based on propositions and the causal model based on clause-length units (Fletcher & Bloom, 1988). Analyses based on different size units appear to be necessary because individual propositions are too small to determine causal relations (e.g., many propositions are single words) and because clause-length units are too large to establish the multilevel hierarchical relationships that are characteristic of the process model.

The differences in the process and the causal models have stimulated recent studies with narrative texts, which compare the models in their predictability of measures of reading performance. As mentioned previously, Trabasso and van den Broek (1985) found that measures derived from the causal model accounted for a substantial proportion of the variance in recall, summarization, and judgments of importance. However, they found that two measures derived from the process model did not account for a significant portion of variance in their data, and the third measure derived from the model had a relationship which was the opposite of that expected. Failure to support the process model in this study may result from the measures they used which were based on argument overlap but did not take into account the processing limitation imposed by short-term memory. O'Brien and Myers (1987) found that measures derived from the causal model predicted a substantial portion of the variance in retrieval times of narrative text, but these data were not predicted by variables derived from the process model. Fletcher and Bloom (1988) found that variables derived from the causal model accounted for a substantial portion of the variance in the recall of narratives. They also found that by combining the process model with the causal model (by using causal relationships to select the items to be retained in short-term memory), the proportion of variance accounted for was greater than for the causal model alone. Thus, in general, these comparative studies show that the causal model has greater predictiveness than the process model; however, both models have some predictiveness when combined.

The present study is also comparative, but with procedural rather than narrative texts. Based on the previous research, it was expected that the causal model would be more predictive of procedural text than the process model. This expectation was also supported by the observation that procedural text appears to have many causal relationships in the form of the steps of the procedures.

Previous research with procedural text has been very limited. Perhaps the study most closely related to the present one was conducted by Graesser (1978). Graesser had people recall texts about common procedures (e.g., how to wash a car or how to catch a fish). He found that subjects' recall of the procedures was predictable from two structural measures that were generated from the answers that an independent group of subjects gave to questions concerning the procedures. The two measures were hierarchical level in the procedure and relational density. The higher level and more densely related statements were recalled better. These findings are intriguing because they demonstrate the effect of structure on recall of procedural text; however, these results may be restricted to procedures with which people are familiar.

Other studies with procedural text have explored the effects of a number of variables, but have not investigated effects of structural variables of the text. Several studies with procedural text have explored the effects of prior knowledge on being able to comprehend the text and on being able to do the task described by the text (Kieras & Bovair, 1986; Kieras, Tibbits, & Bovair, 1984; Mohammed & Swales, 1984). Other studies have manipulated the content of the procedural text to determine what effect that has on performance (Bromage & Mayer, 1981; Dixon, 1987a; Dixon, 1987b; Reder, Charney, & Morgan, 1986). Another group of studies about procedural text has involved manipulating the reader's purpose for reading the text (Dixon, 1987b; Reder, Charney, & Morgan 1986; Schmalhofer & Glavanov, 1986).

In the present research, readers provided importance ratings and recall for procedural texts. These two outcome measures were used to make the research most comparable to previous research. Structural variables derived from the process and causal models were used to predict these data.

EXPERIMENT 1

Method

Each of the eight experimental texts was subjected to propositional and causal analyses to derive the predictor variables. Twenty-four subjects then performed importance ratings of these texts. Subjects performed the ratings three times: once for propositions, once for idea units, and once for sentences.

Texts

The materials were 11 procedural texts (eight experimental and three practice). The criteria for text selection were that no specialized knowledge would be required to understand the text and that the specific task that the text described would be unfamiliar to most people. Texts were selected that appeared to be very different from each other in content and structure. For example, some were primarily a list of procedures while others were well-written text, and some gave reasons for performing the steps, while others did not. The texts were taken directly from naturally occurring instructions, with the exception of "Ohmmeter" which was rewritten to make a list within it seem more text like. The eight experimental texts ranged in length from 115 to 231 propositions (from 34 to 66 phrase-size units). The three practice texts contained approximately 30 propositions each. Table 1 shows the titles and a brief description of these practice and experimental texts. Table 2 shows the texts of "Radio" and "Spool."

Text Analyses

Propositional Analysis

The goal of the propositional analysis was to determine how many times a proposition occurred in working memory during the processing of a text (number of cycles), the number of times a proposition was reinstated (number of reinstatements), and the number of inferences. In addition, the level of the proposition in the hierarchy (level) was determined. The procedures that were followed to derive these measures are described as follows.

First, the texts were coded into propositions using the method described in Bovair and Kieras (1985). Three independent judges coded the texts into propositions and resolved the few discrepancies through discussion. Table 3 shows some of the propositions from the "Radio" text. These particular propositions were selected because they illustrate various aspects of the analysis.

Once the texts were coded into propositions, the second step in the analysis was to "chunk" the propositions using the chunking rules of Miller and Kintsch (1980). Sentences were chunked along sentence boundaries except when the sentences exceeded 19 words, in which case, they were chunked along major phrase boundaries. The propositions in a chunk composed the input propositions for a processing cycle. The dotted lines between the propositions in Table 3 indicate the chunk boundaries.

The next step of the propositional analysis was to build a coherence graph based on the propositions in the chunks, beginning with the input propositions in the first chunk being added to the empty short-term buffer (Miller & Kintsch, 1980). Two judges chose one of these input propositions as the superordinate of the coherence graph. The criterion for superordinate selection was the importance of the information contained in a proposition itself without considering information contained in embedded propositions. Once the superordinate was chosen, the remaining input propositions were connected to the superordinate on the basis of argument overlap, following the guidelines of Miller and Kintsch.

Table 1

Titles and Descriptions of Practice and Experimental Texts

Practice

Bent over row--a short exercise taken from dumbbell instruction book

Upright row--a short exercise taken from dumbbell instruction book

Main burner primary adjustment--one of the adjustments taken from a gas furnace manual

Experimental

How to use a digital timer clock--instructions for the setting and using of and battery replacement for a timer-clock from Radio Shack (clock)[a]

How to fill out employee business expenses form--Part I of Employee business expenses (Form 2106), a federal income tax form (form)

Putting on your protective mask--instructions for using a chemical-biological mask taken from a U.S. Army technical manual (Department of the Army, 1976) (mask)

Wiring and using a shunt-type ohmmeter--based on instructions from the circuit board wiring manual from Science fair 160 in one electronic project kit (Catalog No. 28-258), (Ohmmeter)

Operating a radio set--instructions taken from a U.S. Army operator's manual (Department of the Army, 1980) (radio)

How to make cauliflower and broccoli with buttered rosemary crumbs--recipe taken from Gourmet, November 1987, pp. 162 (recipe)

Performing the torn and restored paper ribbon trick--how to do a magic trick taken from Hay (1975) (ribbon)

How to make and use a spool vehicle--instructions for a children's toy taken from Herbert (1980) (spool)

---

[a]Words in parentheses are shortened titles used in text and tables.

Table 2

Two of the Experimental Texts

Operating a radio set

Operating procedures

To operate radio set AN/PRC-68, perform the following step-by-step procedures.

1. Set the CHAN switch to the proper operating channel. Squad leader will tell you the channel to use.

2. Set the PWR switch to "on." To turn off squelch, turn switch to SQUELCH DIS and hold; it will automatically return to on when you release it.

3. Adjust the VOL control to set the loudness of the received signal.

4. Push the PUSH TO TALK switch and speak into the SPKR/MIC to transmit; release the switch to listen.

5. To turn off radio set AN/PRC-68, set the PWR switch to OFF.

Operation of additional equipment

To operate radio set AN/PRC-68 with a standard handset (H-138/U, H-189/U, or H-250/U), connect the handset to the AUDIO connector and perform the following step-by-step procedures.

1. Line up the keyway (groove) of the handset connector with the keyway of the AUDIO connector and pressing down firmly.

2. Lock the connector by turning it fully clockwise (right). When this is done, the SPKR/MIC is disconnected.

3. Set the PWR switch to ON, and use the push-to-talk switch on the handset to operate.

To operate the radio set AN/PRC-68 at a distance of more than 330 yards or in a poor location (e.g., the foot of a hill), use Antenna AT-892/PRC-25 (long).

1. Disconnect the standard antenna from radio set AN/PRC-68. This is done by turning counterclockwise (left).

2. Install Antenna AT-892/PRC-25 by turning it fully clockwise (right).

Table 2 (continued)

How to make and use a spool vehicle

A very long time ago, someone figured out how to wind up a rubber band inside a spool and then make the rubber band unwind slowly to make the spool crawl across the floor.

A modern version uses a spool of any size. Through the opening, slip a rubber band, which is about the same length as the spool, so that it passes from one end of the spool to the other. Anchor one end of the rubber band with a tack to the end of the spool. Slip the other end of the rubber band through the hole in a metal washer. The washer creates enough friction to keep the rubber band from unwinding quickly and yet is slippery enough to allow the wheel to turn slowly. That's why the washer is called a slipper. If you don't have a washer, you can make a slipper from the plastic of a refrigerator container or coffee-can top.

Into the end of the rubber band and through the slipper, put the last part--the drag. It can be a match stick, knitting needle, pencil, and so on.

If the rubber band is not taut, wind it several times around the drag to take up the slack.

To energize your spool vehicle, turn the drag to wind up the rubber band inside the spool. Set the spool down on the table. The drag keeps its end of the rubber band from turning. The twisting action of the rubber band is transferred via the anchor to the spool. With the right amount of friction from the slipper, the spool continues across the table until most of the energy is released.

You can have races and battles if your friends also make spool vehicles.

Table 3

Propositions From the Radio Text

Operating a radio set

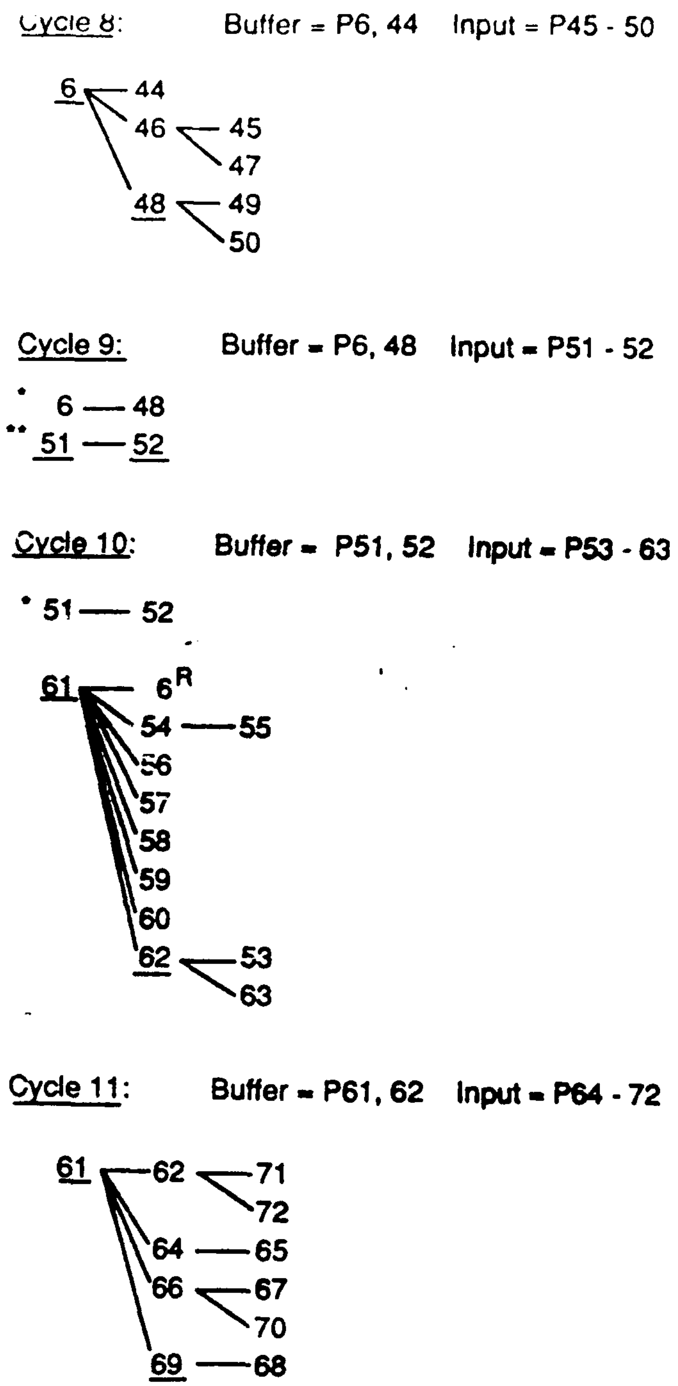
```
 .
 .
 .
 6  (perform $ procedure)
 .
 .
 .
44  (listen $)
    ----------
45  (in-order-to P50 P46)
46  (turn-off $ radio-set)
47  (label radio-set AN/PRC-68)
48  (set $ switch2)
49  (label switch2 PWR)
50  (to P48 OFF)
----------
51  (of operation equipment)
52  (mod equipment additional)
----------
53  (in-order-to P62 P56)
54  (operate $ radio-set)
55  (label radio-set AN/PRC-68)
56  (with P54 handset)
57  (mod handset standard)
58  (can-be handset H-138/U)
59  (can-be handset H-189/U)
60  (can-be handset H-250/U)
61  (connect $ handset)
62  (to P61 connector1)
63  (label connector1 AUDIO)
----------
64  (do $ P62)
65  (by P64 P73)
66  (line-up $ keyway1)
67  (isa keyway1 groove)
68  (possess connector2 keyway1)
69  (mod connector2 handset)
70  (with P66 keyway2)
71  (possess connector1 keyway2)
 .
 .
 .
```

Starting at the first proposition in a processing cycle, the propositions were read from the top to the bottom of that input cycle, and propositions that contained an argument in common with the superordinate were connected to the first level of the coherence graph. The placeholder "$", which was used to represent an unspecified text argument (e.g., you) in a proposition, was counted as an argument when connecting propositions. When no more propositions could be connected to the first level, the list of unconnected propositional inputs were reread in their natural order (from top to bottom) and were tested for argument overlap, beginning at the bottom of the second level of the coherence graph and subsequently moving up through all the connected propositions at the second level (i.e., the recency rule was used). If more than one unconnected proposition overlapped with an already connected proposition, they were connected in numerical order with propositions occurring later connected below earlier ones.

After all the propositions in an input chunk were connected, the buffer propositions (i.e., those propositions that would remain in short-term memory and that would serve as connectors for the input propositions for the next input cycle) were chosen according to the modified leading edge strategy (Miller & Kintsch, 1980). This technique first involved choosing the superordinate proposition and any propositions embedded in it. Then, if the number of propositions did not exceed the buffer size (**s**), propositions that fell along the bottom of each level (and their embedded propositions) were chosen for the buffer. The process stopped when **s** propositions had been selected. Miller and Kintsch found that restricting the value of **s** to 1 or 2 resulted in the best fit for their recall data. Spilich, Vesonder, Chiesi, and Voss (1979) also found an **s** of 2 provided the best fit. Therefore the **s** was set at 2 for the propositional analyses. The buffer size was **s** + 1 for the first cycle because the buffer was empty at the beginning of the text. Any cycle could stretch its buffer by 1 if a selected proposition contained an embedded proposition.

Then the input propositions from the second cycle were connected to the buffer propositions and to each other on the basis of argument overlap. Once all the propositions in the second cycle had been connected, the buffer propositions for the third cycle were chosen, and the process was repeated for this cycle and the remaining ones. If a connection could not be made between the propositions in an input cycle and the buffer, a search was begun for a reinstatement (i.e., a previously occurring proposition that contained an argument which would overlap with a currently active proposition). Previous cycles were searched in order of recency, and within a cycle, the search proceeded in the order of recency from the highest to the lowest level. If a reinstatement was found, it was put at the beginning of the current input list. The new subgraph was begun with a new superordinate, selected from the current input list, and the remaining propositions were connected according to argument overlap. If no reinstatement could be found, an inference was made and the remaining proposition(s) was connected. Figure 1 shows part of the coherence graph for "Radio" based on the propositions in Table 3. Note this starts in the middle of the text with Cycle 8.

When the coherence graph was complete (i.e., all the propositions had been connected), the judges counted the number of cycles in which each proposition occurred. If a proposition was never placed in a buffer or reinstated, it occurred once; if a proposition was placed in the buffer twice and was reinstated once, it occurred four times. The judges also determined the number of reinstatements for each proposition, as well as the number of inferences. The level of the proposition in the hierarchy was determined by



Note. R - Reinstatements
* - No connection with input
** - Inference required

Figure 1. Part of the coherence graph for the radio text.

the coherence graph (Kintsch & Keenan, 1973). The first level propositions were the superordinates and the second level propositions were those connected to the superordinates. Third, fourth, and subsequent level propositions were determined in the same manner. If a proposition occurred at two different levels in the hierarchy (e.g., a reinstatement), the more superordinate level it held was considered to be its level.

Causal Analyses

Causal analyses were performed to determine how many causal relationships each element of the text had with other text elements (total connections) and whether an element was on the causal chain which led from the opening to the closing of the text (Causal Chain). These analyses were performed using the following procedures.

The experimental texts were first parsed into idea units (a sentence or part of a sentence that expresses a complete idea which contains an actual or implied verb and is usually a phrase-size unit). Four independent judges parsed one of the texts ("Clock") together. Then, they parsed the remaining seven texts independently. Judgment discrepancies were resolved through discussion. Agreement for each text was determined by dividing each judge's correct number of idea units by the total number of agreed upon units. An average agreement across the four judges was then determined for each text, ranging from 83% on the "Form" text to 96% on the "Radio" text. The average agreement across these seven texts was 91%. Table 4 shows part of the idea units from the "Radio" text.

Table 4

Idea Units From the Radio Text

Idea unit	Radio text
22	To operate Radio Set AN/PRC-68 with a standard handset (H-138/U, H-189/U, or H-250/U),
23	Connect the handset to the AUDIO connector.
24	This is done
25	By lining up the keyway (groove) of the handset connector with the keyway of the AUDIO connector
26	And pressing down firmly.
27	Lock the connector
28	By turning it fully clockwise (right).
29	When this is done,
30	The SPKR/MIC is disconnected.
31	Set the PWR switch to ON,
32	And use the push-to-talk switch on the handset
33	To operate.

The causal analyses were based on the idea units, following the guidelines of Trabasso and Sperry (1985). Three independent judges considered whether each idea was causally related to each idea unit in a particular text. For one unit to be causally related to another, it had to pass several tests. First, the two units had to pass the counterfactual reasoning test which is of the form "If not A, then not B." For example, the two idea units "adjust the volume control" and "to set the loudness of the received signal" passed the counterfactual reasoning test because if one did not adjust the volume control, one could not set the loudness of the received signal. Whenever two units were subjected to this test, the context of the statements was taken into account. Often two units that would not have been judged as causally related outside the current situation were found to have a causal relationship "in the circumstances." For example, in "Recipe," the two units "transfer the crumbs to a bowl" and "stir the herbs into the crumbs" were related in the context of the recipe because the crumbs had to be removed from the skillet before the herbs were added; otherwise, the herbs would have been cooked.

Once a pair of idea units passed the counterfactual test, the next criterion was temporal precedence. The idea unit that caused another unit had to occur temporally before it (but not necessarily before in the text). The final criterion was directness. Two units were judged to be directly related if no other idea unit mediated the two units. If the relationship was judged to be indirect, the two units were not causally related.

Another possible type of relationship between two units (in addition to causal) was that of temporal coexistence. Generally, idea units were judged to coexist if they occurred at the same time. Temporal coexistence relationships also occurred when background or setting information was being provided, when an idea was restated, and when alternatives were being presented.

Once the causal analysis of a text was completed and the discrepancies were resolved through discussion, a diagram of the causal connections was made. Figure 2 shows part of the causal diagram for "Radio." An elaboration of the relationships between these units is provided in the Appendix.

The causal diagrams were used to determine the number of each kind of connection an idea unit had with other idea units. Arrows leading into a unit indicated that the unit was being caused by another unit (ins); arrows leading out of a unit indicated that that unit was causing another unit (outs). Intersection signs indicated temporal coexistence between two units (intersections). To determine the total connections for a particular unit, the judges added the ins, outs, and intersections as specified by Trabasso and Sperry (1985).

The diagrams were also used to determine which of the idea units were members of the causal chain. The chain for procedural text starts with the introduction of the task and ends with its completion. The beginning and end of the causal chains for each text were determined by three judges. Then the causal chains were determined independently by two judges and were checked against each other for errors. An idea unit was judged to be a member of the causal chain if it had antecedents that led back to the chain's opening and

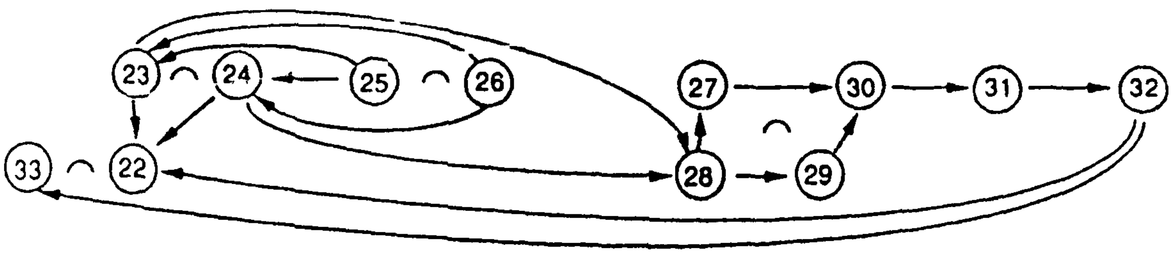


Figure 2. Part of the causal diagram relationships for the radio text.

consequences that led forward to its closing. Idea units that are members of the causal chain are circled in Figure 2. In the illustrated section of the causal diagram, all the units are on the causal chain. However, this is not typical; the proportion of idea units that were on the causal chain ranged from .23 (mask) through .85 (recipe), with a mean of .57 over all the texts. The causal chain variable had values of either 1 or 0 depending on whether the unit was a member of the chain.

Three judges performed the causal analysis twice on each of the experimental texts. Each time, discrepancies were resolved through discussion. The first set of analyses was performed during the course of several months. To ensure that consistent criteria were used, the original analyses of the eight texts were rechecked within 10 days.

To check interjudge agreement for the causal analyses, Kappas were calculated (Cohen, 1960) for three of the experimental texts. Agreement of each judge with each other judge was determined for the second set of analyses upon which the regression analyses were based. The Kappas may be inflated because this second set of analyses was based on the first set. However, in all cases, all judges made a substantial number of changes in the second set of analyses. The $\underline{K}$ statistic (Cohen, 1960) ranged from .73 through .87, with an average $\underline{K}$ of .80. Percent agreement (number of actual agreements/number of possible agreements x 100) ranged from 96% through 98%, with an average of 97%.

Ratings

Subjects

The subjects were 24 female college students from a liberal arts college and were paid $5.00 per hour. Their time to complete the required tasks ranged from 4.3 hours through 8.4 hours with a mean time of 5.9 hours. Two additional subjects never returned after the initial training session and were replaced.

Materials

Each of the three practice and the eight experimental texts was prepared in three different sized units: sentences, idea units, and propositions. In all cases, the response sheets were prepared with the same format: the text title appeared at the top of the first page, followed by the text in standard paragraph format. A 7-point rating scale appeared after the text and at the top of all the subsequent pages. The scale was labeled with the number 1 corresponding to "unimportant to task performance," the number 4 corresponding to "useful but not essential to task performance," and the number 7 corresponding to "essential to task performance." The title appeared again under the first occurrence of the rating scale, followed by the text broken into units of the appropriate size. Each unit was typed on a separate line, with a blank space for the rating response appearing on the same line to the left of the unit. If necessary, a second line of the unit appeared on the next line without a blank. The text units were presented in the same order as they occurred in the text.

When propositions were presented on the rating form, normal English was used ([Add the crumbs] rather than the proposition format [add $ crumb]). If a proposition consisted of a modification (e.g., [mod crumb buttered]), the modifier was presented in brackets in the position it would normally hold (add the [buttered] crumbs) and then presented separately inside slashes (/buttered/) on the next line for subjects' rating. Subjects were instructed to ignore the bracketed words when making their judgments of importance.

Procedure

Subjects were divided into three groups of eight subjects and each attended three training sessions which were 1 week apart. Each training session was for a different text unit type (i.e., sentences, idea units, and propositions). All groups rated all three text unit types but in different orders. The orders were arranged so that each unit type occurred in each of the three serial positions. When a subject could not attend a group training, she received training individually.

Subjects were first given an introduction to the purpose of the study. Subjects were told to make their ratings based on the importance of the unit to the overall procedure described in the text and that there were no right or wrong answers. Subjects were asked to read through each text completely before making any ratings for that text. The practice rating forms for the appropriate unit type were distributed and the experimenter reviewed the first practice text and explained her reasons for making the ratings. After the experimenter rated the first practice text, the subjects were instructed to read the next two practice texts and rate the units. They rated the second and third practice texts at their own pace. When they finished, they turned in their practice rating forms and received a packet which consisted of the eight experimental texts, all the same unit type.

Each subject in a group rated the texts in a different order. The eight different orders for presenting the text were based on a Latin square. The subjects were asked to gauge their time so that they always finished a text before leaving a session and to rate the texts in the order they appeared in the booklet. Subjects returned during the week at scheduled times to rate the remaining experimental texts for that week. This same general procedure was used in the following 2 weeks, except that each week, a given subject rated the texts in a different order. By the end of the third week, each subject had rated all eight texts presented in all three unit types.

Results

Means versus Medians as Outcome Measures

The calculated mean and median importance ratings were used in separate analyses as outcome measures. Preliminary analyses indicated that the results were very similar; therefore, only analyses for mean importance ratings are reported.

Reliability of Ratings

For each size unit, the mean rating was calculated for each text unit for 12 subjects (half of those subjects who rated propositions first, half of those who rated idea units first, and half of those who rated sentences first). The mean rating for each text unit was then calculated for the remaining 12 subjects.

Correlations were calculated between the mean ratings for the individual units for the two groups of 12 subjects. Separate correlations were performed for each of the eight texts for each size unit, resulting in 24 correlations. The values of the correlations ranged from .84 through .99. The mean reliability coefficients for the propositions, idea units, and sentences were .92 (N=115 to 230), .92 (N=35 to 66), and .95 (N=17 to 28), respectively. The mean ratings for the individual units ranged from about 3 through 7 for each size unit (e.g., for "clock" the range of mean ratings was 3.17-7.00 for propositions, 3.83-7.00 for idea units, and 3.92-7.00 for sentences). The reliability data, together with the response variability, suggest that all three unit types were rated meaningfully and consistently by the subjects.

Propositional Analyses

Three predictors were used in the regression analyses for the proposition ratings: Number of cycles, number of reinstatements, and level. Number of inferences were not used as a predictor because there were only two inferences in all the texts.

The authors performed regression analyses of each of the texts, forcing each of the predictor variables to enter first and last. The proportion of variance that each predictor shared with mean importance rating when none of the other variables were entered is called "alone" in Table 5; the proportion of variance that each predictor shared with mean importance rating after all the other predictors had been entered is called "unique." Table 5 shows the alone, unique, total, and adjusted R^2 for each text and the mean R^2 across all the texts. Half of the texts do not contain a row labeled reinstatements because these texts did not contain reinstatements.

Number of cycles accounted for a significant proportion of the variance in importance rating in two of the texts when it was entered first and in one of the texts when it was entered last. Number of reinstatements did not make a significant contribution to any of the texts when entered either first or last. Level accounted for a significant proportion of the variance in importance rating in six of the texts when it was entered first, and in three texts when it was entered last. The total R^2 was significant for five of the texts. The mean total R^2 across the eight texts was .043 ($p<.001$) and based on the mean R^2 values both number of cycles and level accounted for a significant portion of variance when entered either first or last.

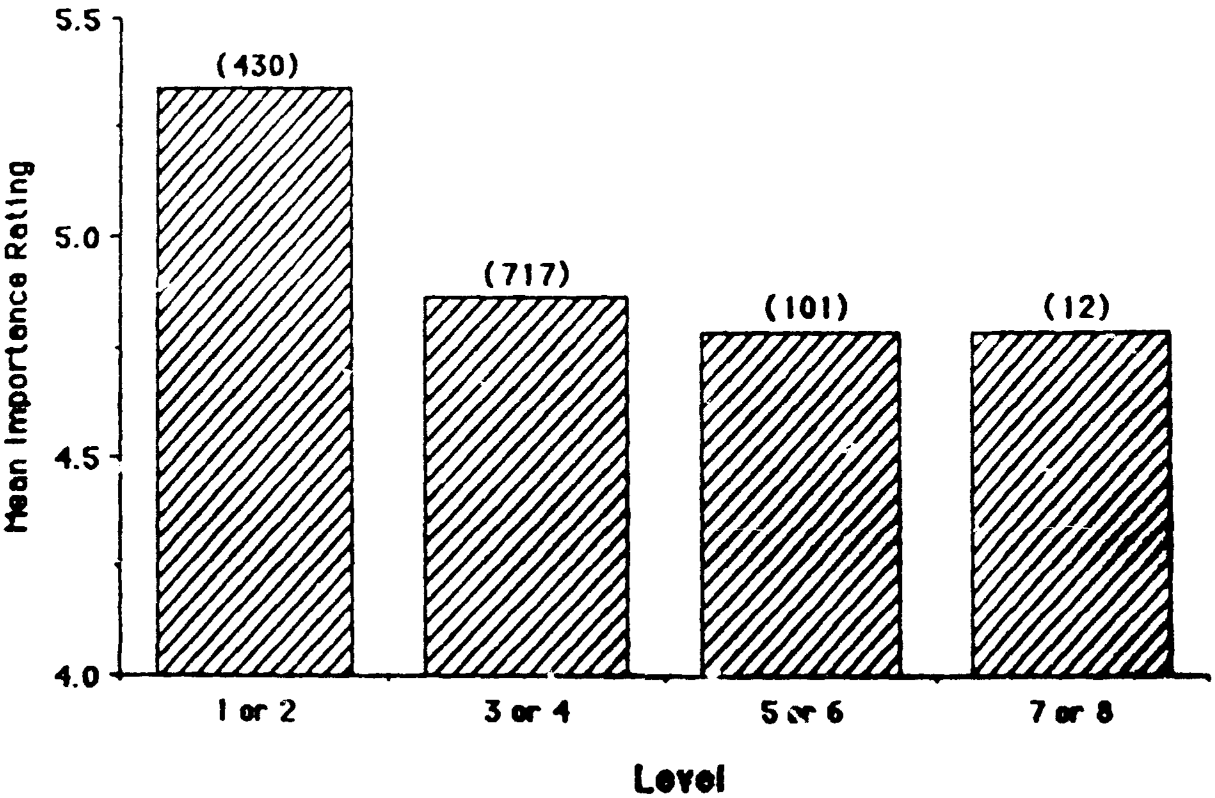
An analysis of variance (ANOVA) was performed on the mean importance ratings of propositions for all combined texts with level and number of cycles as factors. For this analysis, adjacent values of each variable were combined because of the small number of cases at the higher levels. The results showed that level was significant, $F(3,1252)=16.26$, $p<.001$. The number of cycles and the interaction were not significant. As shown in Figure 3, as level increases (i.e., a proposition becomes more subordinate), it is rated as less important. A post hoc contrast showed that level 1 or 2 was significantly different from levels 3 or 4 and from levels 5 or 6 ($p<.001$). Other contrasts were not significant. The number of cycles did not reach significance, but there was a trend that suggested that as the number of cycles increased mean importance ratings also increased.

Table 5

Proportion of Variance Accounted for in the Importance Ratings by the Predictors Derived From the Process Model

Text	Total R^2	(Adjusted R^2)	Alone	Unique
Clock	.104**	(.080)		
		Cycles	.089***	.042*
		Reinstatements	.007	.001
		Level	.057**	.014
Form	.034	(.022)		
		Cycles	.014	.007
		Level	.027*	.020
Mask	.042*	(.030)		
		Cycles	.013	.005
		Level	.037*	.029*
Ohmmeter	.042*	(.026)		
		Cycles	.021*	.008
		Reinstatements	.008	.005
		Level	.028*	.015
Radio	.067*	(.043)		
		Cycles	.013	.001
		Reinstatements	.003	.001
		Level	.066**	.051*
Recipe	.085**	(.074)		
		Cycles	.019	.002
		Level	.084***	.066***
Ribbon	.014	(.005)		
		Cycles	.000	.000
		Level	.014	.014
Spool	.027	(.002)		
		Cycles	.005	.001
		Reinstatements	.010	.014
		Level	.013	.015
Mean of all texts	.043***	(.025)		
		Cycles	.012***	.004*
		Reinstatements	.006	.004
		Level	.035***	.025***

*$p < .05$. **$p < .01$. ***$p < .001$.



Note. Numbers in parentheses indicate the number of propositions.

Figure 3. Mean importance ratings of propositions as a function of level in Experiment 1.

While the number of cycles and the level reached significance in some of the texts and analyses, the proportion of variance in importance ratings for which they accounted was not very large. Thus, it appears that the measures derived from the propositional analysis are not powerful predictors of importance ratings for procedural texts.

Causal Analyses

Total connections and causal chain were used as predictors of importance ratings in the regression analyses for the idea unit ratings. As with the propositional analyses, a regression analysis was performed of each of the eight texts, and each predictor was forced to enter first and last. An interaction term, Total Connections x Causal Chain, was entered between the causal variables (Trabasso & Sperry, 1985). Therefore, the unique in Table 6 indicates the proportion of variance in mean importance ratings--a predictor shared after both the other predictor and the interaction term had been entered.

Table 6 shows alone, unique, total, and adjusted R^2 for each text and the mean R^2 across all the texts. Total connections was a significant predictor in two of the texts when it was entered first. Causal chain accounted for a significant proportion of the variance in importance ratings in five of the texts when it was entered first. Neither of the variables reached significance when entered last. When total connections was entered first, the interaction term was significant ($p<.05$) in five texts ("clock," "form," " mask," "ribbon," and "spool"). When causal chain was entered first, the interaction term was significant in two texts ("clock" and "form"). The total R^2 was significant for five of the texts. The mean total R^2 across the eight texts was .158 ($p<.001$). The mean R^2 across the texts was significant for both total connections and causal chain when entered either first or last.

The relationship between mean importance rating and total connections is shown in Figure 4. Total connections is broken down by idea units which were on the causal chain and those which were off the chain. An ANOVA was performed using the mean importance rating of idea units for all combined texts with total connections and causal chain as factors. For this analysis, adjacent levels of the total connections variable were combined because of the small number of cases for the larger total connections, and total connections of six and more were excluded from the analysis since none of these were off the chain. This analysis revealed a significant effect of causal chain, $F(1,355)$ [illegible]7.10, $p<.001$. Total connections and the interaction were not significant. As expected, ratings were higher for the idea units which were on the causal chain.

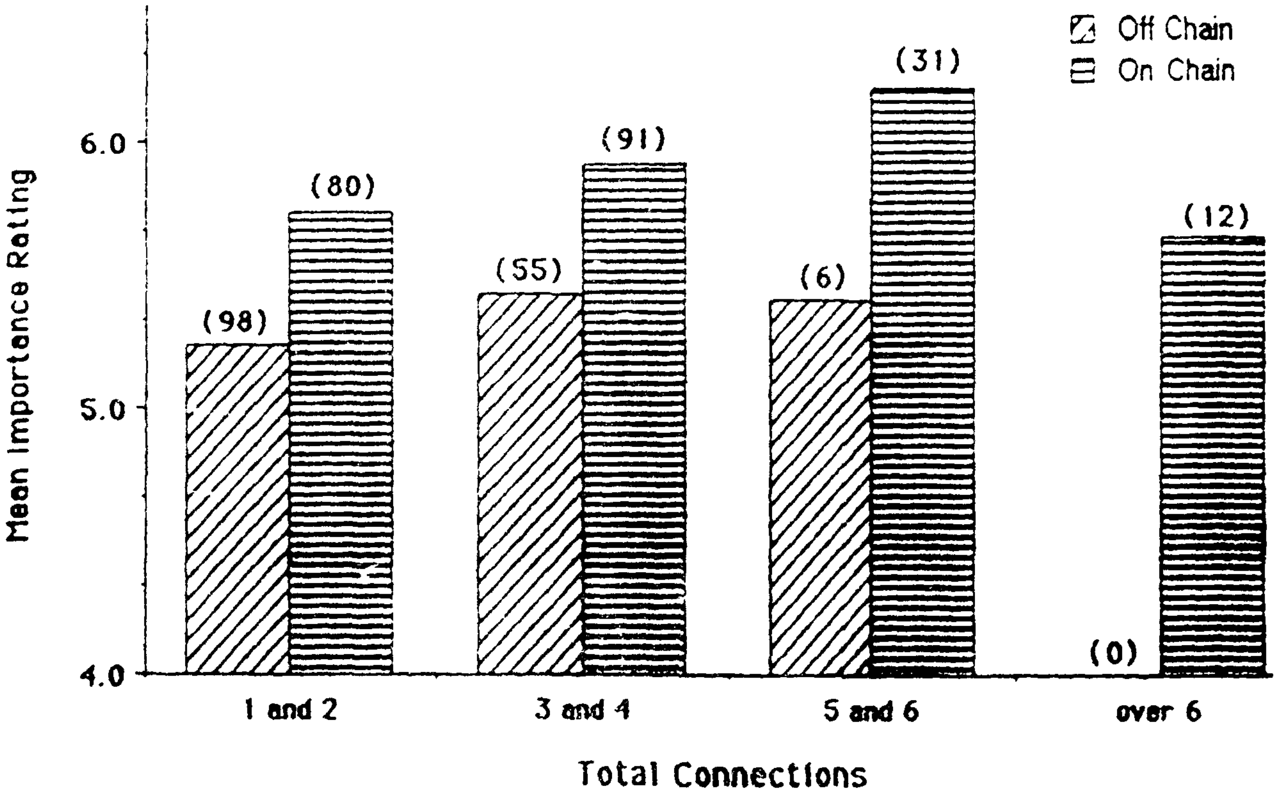
While total connections did not reach significance ($p<.08$), generally as the number of connections increased, the importance rating increased. The exception is when the number of connections is greater than six (not included in this analysis). An inspection of these idea units showed that two types of units appeared to have more than six connections, units essential for performance and those not essential (e.g., highly redundant information). The five units essential had a mean of 4.82. This suggests that some idea units may have many connections but not be important to the task. While not reflected in the ANOVA of all texts combined, total connections appears to have a greater effect for the units on the chain than for those off the chain, which is suggested by the significant interactions obtained in the regression analyses of the individual texts.

Table 6

Proportion of Variance Accounted for in the Importance Ratings by the Predictors Derived From the Causal Model

Text	Total R^2	(Adjusted R^2)	Alone	Unique
Clock	.171*	(.115)		
		Total connections	.088*	.070
		Causal chain	.016	.008
Form	.245*	(.183)		
		Total connections	.001	.036
		Causal chain	.103*	.082
Mask	.176*	(.120)		
		Total connections	.051	.011
		Causal chain	.162**	.013
Ohmmeter	.042	(-.026)		
		Total connections	.010	.000
		Causal chain	.042	.005
Radio	.135	(.065)		
		Total connections	.003	.001
		Causal chain	.122*	.041
Recipe	.087	(.024)		
		Total connections	.036	.012
		Causal chain	.035	.009
Ribbon	.191**	(.152)		
		Total connections	.033	.008
		Causal chain	.174**	.002
Spool	.323**	(.255)		
		Total connections	.215**	.003
		Causal chain	.272**	.006
Mean of all texts	.158***	(.085)		
		Total connections	.035***	.012*
		Causal chain	.101***	.014*

*$p < .05$. **$p < .01$. ***$p < .001$.



Note. Numbers in parentheses indicate the number of idea units.

Figure 4. Mean ratings of idea units as a function of total number of connections and presence on or off causal chain in Experiment 1.

The analyses showed that total connections was not a particularly good predictor of the ratings; however, one of its components (ins, outs, and interactions) could be. To test this, additional regression analyses were performed using the data for each text to determine whether the kind of connection an idea unit had with another idea unit was important. For these analyses, ins, outs, and intersections were used in addition to causal chain. The results of these analyses showed that when ins was entered alone as the sole predictor of ratings, it was significant in two of the texts ($p<.05$). Outs was a significant predictor in four texts ($p<.05$), and intersection was significant in only one text ($p<.05$). These analyses showed that outs was significant in more texts than ins, intersections or total connections. However, as in other analyses, the size of the R^2 values was not very large. Thus, the different kinds of connection did not serve as much better predictors than did total connections.

Another interesting finding that can be seen in Figure 4 concerns the frequency of idea units that occurred as a function of total connections and causal chain. A chi-square analysis performed using these frequencies revealed a significant effect ($\chi^2(3)=32.19$, $p<.01$). As can be seen, with more total connections, proportionally more idea units were on the Causal Chain, which is what would be expected.

While the proportion of variance in importance ratings of idea units accounted for by total connections and causal chain is not overwhelming, it is substantial in some of the texts. Causal chain was significant in more texts than total connections. The total R^2 for the causal model for all texts combined was larger than for the process model ($p<.001$).

Sentence Analyses

Because the propositions are generally smaller than the idea units, it is possible that any differences in the proportion of variance accounted for by the process and causal models are because of some factors other than the goodness of these models (e.g., the ease with which the units were rated). The importance ratings of sentences were analyzed to determine whether the same patterns of results were obtained when the process and causal predictors were used to predict the ratings of a common unit, the sentence.

Propositions and idea units were collapsed within each sentence and sentence means were determined for each of the relevant predictors (i.e., number of cycles, number of reinstatements, and level for the process model, total connections and causal chain for the causal model). The mean sentence importance rating was used as the outcome measure. The authors performed the same type of regression analyses with the sentence units as with the propositions and idea units described previously (i.e., each of the predictor variables was forced to enter first and last). In the causal regression, the interaction term, Total Connections x Causal Chain, was entered between the predictor variables.

The mean total R^2 across all the texts was computed from the process and causal model analyses for the sentence ratings. The same pattern of results emerged (i.e., the causal model accounted for a larger overall proportion of variance) in importance ratings (Mean R^2 = .231 [$p<.001$], range of R^2 = .013 through .459 for individual texts with three texts having $p<.05$) than did the process model (Mean R^2 = .119 [$p<.001$], range of R^2 = .018 to .360 for individual texts with no text reaching significance). However, for these data, the difference between the mean R^2 for the process and for the causal was not significant. Therefore, it appears that the greater proportion of variance accounted for by the causal model cannot be explained solely on the basis of some different characteristics of the ratings for the different size units.

EXPERIMENT 2

The results in Experiment 1 showed that the causal model of Trabasso and his colleagues (1984, 1985) predicted the importance ratings for the idea units better than the process model of Kintsch and his associates (1973, 1978, 1985) predicted the importance ratings for the propositions. The causal model also accounted for more variance in importance ratings for sentences than the process model. Thus, in agreement with the previous comparative studies performed with narratives, these results suggest that the causal model may be more appropriate for procedural texts than the process model.

However, one of the reasons that the process variables may have been poorer predictors in this study than expected based on previous research (Miller & Kintsch, 1980), is that this study used importance ratings as the dependent measure instead of recall. Importance ratings do not appear to have been studied by Kintsch and his colleagues. While ratings and recall are highly correlated in previous research (Brown & Smiley, 1977), it may be that the predictors from the causal model are better for the importance ratings but those from the process model are better for recall. As a result, this study conducted a second experiment with the same texts using recall as the dependent measure.

Method

Sixteen subjects read and recalled each of the eight texts in a different order.

Subjects

The subjects were 16 female college students from a liberal arts college. They were either paid at the rate of $5.00 per hour or received experimental credit in their psychology class. Their time to complete the required tasks ranged from 2.5 through 3.8 hours with a mean of 3.0 hours. One additional subject failed to return after the initial session, and her data were not included in the analyses.

Texts and Materials

Eight experimental texts and two practice texts from Experiment 1 were used. They were arranged in two booklets for each subject. Each booklet contained one of the practice texts followed by four of the experimental texts. The experimental texts were arranged in a different order for each subject based on a Latin square. Following each text were two pieces of blank paper for the recall.

Procedure

Subjects were run in a group with each subject attending two sessions which were a week apart. For three subjects who could not attend one of the group sessions, individual sessions were arranged.

In the initial session, subjects were given their first recall booklet and an introduction to the purpose of the study. They were told to read each text twice at their normal reading speed (Birkmire, 1985). Subjects were instructed to read the text twice to ensure a reasonable level of recall and less random variability in recall. They were advised to try to understand the text and not to try to memorize it. After reading a text, subjects were instructed to turn to the next page in their booklets and write down as much of the text as they could using the exact words when possible or paraphrasing it when they could not remember the exact words.

Then, subjects read and recalled the practice text and then the four experimental texts in the first booklet one at a time. During the second session, subjects were reminded of the purpose of the study, and then they read and recalled the remaining texts in the second booklet.

Protocol Scoring

Recall protocols were scored in two ways. One way was based on propositions as the unit of analysis, and the other was based on idea units. A lenient set of criteria was developed for scoring each size unit in which a given unit was counted correct if the meaning was retained. Lenient criteria were used to make the proposition and idea unit scoring more comparable. For both ways of scoring the recall protocols, the unit was counted as correct only if the information was included in the recall protocol (i.e., inferences were not made). If information from two units was combined during recall, both were counted as correct. The units did not have to be recalled in the order given as in the text unless context was needed to clarify which unit was being recalled. For both ways of scoring the protocols, if a word was consistently substituted in the recall protocol for another (e.g., helmet versus mask), credit was given for the unit. For idea units, if the main idea (the object and verb) was correct but some part of the unit was omitted (e.g., a prepositional phrase) or an adjective was incorrect (e.g. left versus right), the unit was counted as correct.

Two judges judged all eight texts for propositions and two other judges judged the idea units. The pairs of judges resolved discrepancies through discussion. In the few cases when the judges could not resolve the discrepancies, a third judge did so. The mean percentage of agreement between judges for the propositions was 98% with a range of 97% through 99% for individual texts, and the mean percentage agreement was 91% for the idea units with a range of 89% through 94% for individual texts.

After all the recall protocols were scored, subjects who got each proposition and idea unit correct were determined and those subjects served as the recall measure in the following regression analyses.

RESULTS

Recall Scores

For propositions over all the texts, 28.7% of the subjects recalled each unit. The percentage recall varied from 18.5% for the "ribbon" text through 41.9% for the "clock" text. For idea units during all texts, the mean percentage of subjects who recalled each idea unit was 39.1%. The average percentage varied from 28.9% for the "ohmmeter" text through 58.1% for the "clock" text. There was a wide range of recall for the individual propositions and idea units in a given text. For most of the text, the range was between 0% through 100% recall for both units; however, for some texts, there was a somewhat smaller range (e.g., for spool, the range was 6% through 93% for the idea units). Thus, for each of the texts, some of the units had very low levels of recall, and other units had very high levels.

Propositional Analyses

The propositions recalled were analyzed in regression analyses as done for importance ratings with predictors derived from the process model. The results of these analyses are shown in Table 7. As shown by comparing Table 5 with Table 7, the R^2 values based on recall are generally smaller than those based on importance ratings with three (rather than five) significant total R^2 values for the recall data.

An ANOVA was performed using the recall measures for the propositions for all combined texts with number of cycles and level as factors. For this analysis, adjacent values of each variable were combined. The results showed that number of cycles [$F(3,1252)=4.38, p<.01$] and level [$F(3,1252)=4.06, p<.01$] were both significant. The interaction was not significant. As shown in Figure 5, as the number of cycles increased, there was a tendency for recall to increase.

Post hoc contrasts showed that level 1 or 2 differed from level 5 or 6 ($p<.01$). However, these results should be interpreted cautiously because of the small N for levels 5 and 6 cycles. The relationship between recall and level is shown in Figure 6. Unexpectedly, as level increased, a proposition was recalled better. Post hoc contrasts showed that both level 1 or 2 and Level 3 or 4 were significantly different from level 5 or 6 ($p<.01$).

It appears that for procedural text, the predictors from the process model are not better for recall than for importance ratings. This supports the conclusion of Experiment 1 that the process model is not a powerful model for procedural text.

Causal Analyses

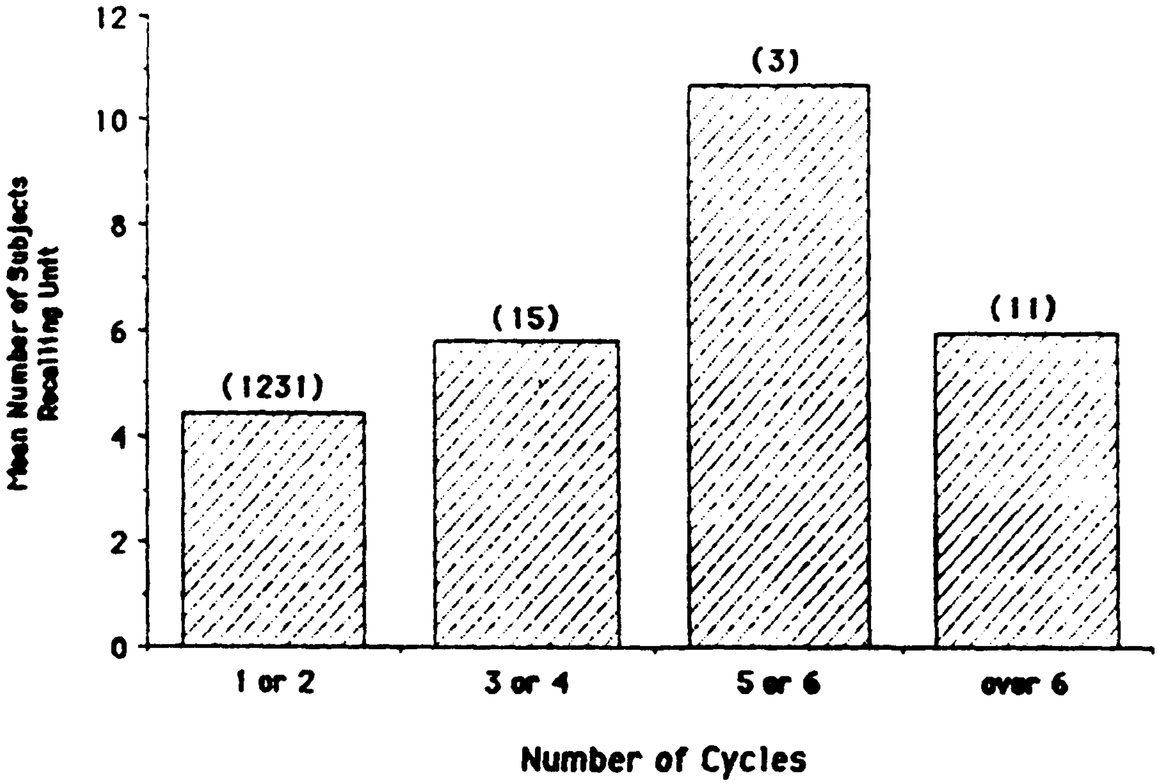
The idea unit recall measure was analyzed using regression analyses like those used for importance ratings with predictors derived from the causal model. These results are shown in Table 8. As shown by comparing Table 6 with Table 8, the R^2 values based on recall are generally smaller than those based on importance ratings with one (rather than five) significant total R^2 values for the recall data. The relationship between the recall measure and total connections and causal chain were analyzed in an ANOVA for all texts

Table 7

Proportion of Variance Accounted for in the Recall for Each Text by the Predictors Derived From the Process Model

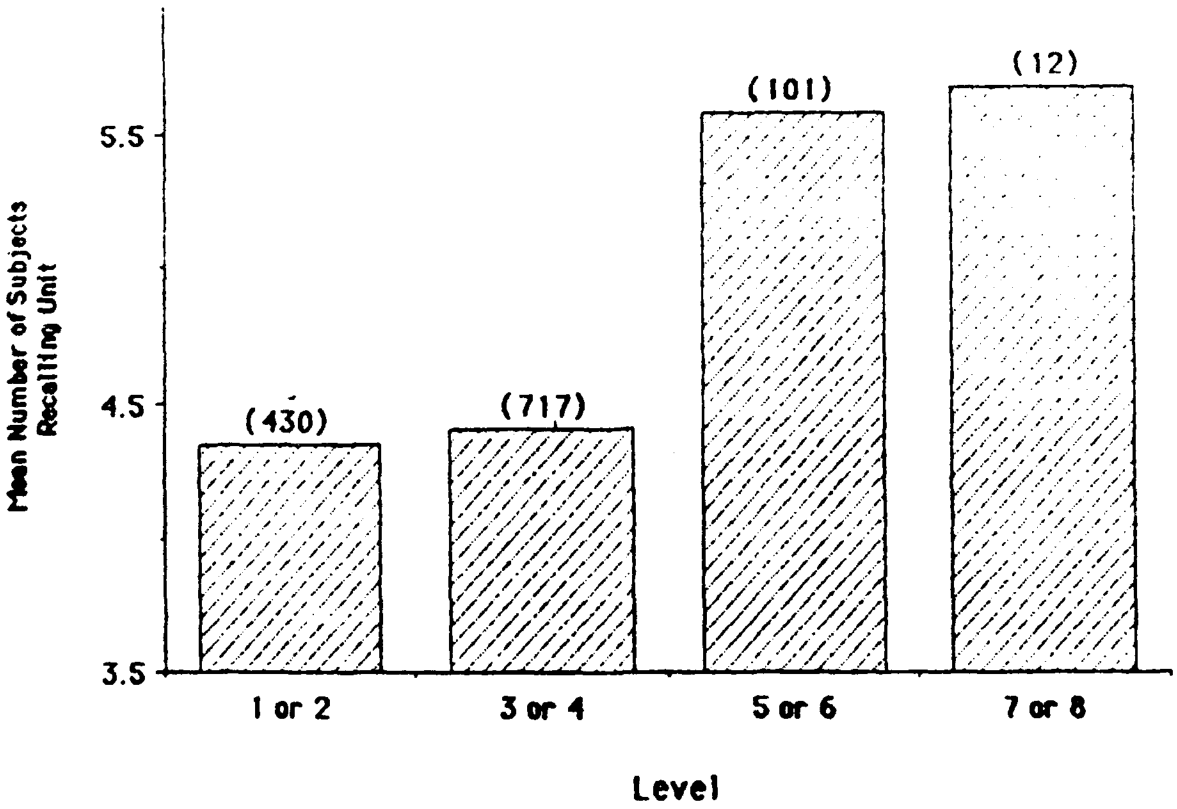
Text	Total R^2	(Adjusted R^2)	Alone	Unique
Clock	.112**	(.008)		
		Cycles	.057*	.012
		Reinstatement	.000	.000
		Level	.100***	.054*
Form	.017	(.004)		
		Cycles	.015	.017
		Level	.000	.001
Mask	.058**	(.046)		
		Cycles	.056**	.049**
		Level	.009	.002
Ohmmeter	.032	(.016)		
		Cycles	.005	.000
		Reinstatements	.000	.000
		Level	.031*	.026*
Radio	.009	(-.017)		
		Cycles	.007	.000
		Reinstatements	.004	.000
		Level	.006	.002
Recipe	.021	(.003)		
		Cycles	.008	.001
		Level	.019	.013
Ribbon	.067***	(.059)		
		Cycles	.061***	.052***
		Level	.015	.006
Spool	.014	(-.010)		
		Cycles	.002	.007
		Reinstatements	.003	.000
		Level	.005	.010
Mean of all texts	.037***	(.019)		
		Cycles	.022***	.012***
		Reinstatements	.000	.000
		Level	.015***	.010***

*$p < .05$. **$p < .01$. ***$p < .001$.



Note. Numbers in parentheses indicate the number of cases.

Figure 5. Mean number of subjects recalling propositions as a function of number of cycles in Experiment 2.



Note. Numbers in parentheses indicate the number of cases.

Figure 6. Mean number of subjects recalling propositions as a function of level in Experiment 2.

Table 8

Proportion of Variance Accounted for in the Recall for Each Text by the Predictors Derived From the Causal Model

Text	Total R^2	(Adjusted R^2)	Alone	Unique
Clock	.069	(.007)		
		Total connections	.008	.021
		Causal chain	.044	.008
Form	.216*	(.153)		
		Total connections	.022	.113*
		Causal chain	.100*	.008
Mask	.088	(.026)		
		Total connections	.001	.001
		Causal chain	.062	.053
Ohmmeter	.120	(.057)		
		Total connections	.035	.071
		Causal chain	.010	.002
Radio	.135	(.065)		
		Total connections	.083	.006
		Causal chain	.106*	.010
Recipe	.041	(-.025)		
		Total connections	.024	.001
		Causal chain	.016	.001
Ribbon	.047	(.001)		
		Total connections	.001	.033
		Causal chain	.000	.038
Spool	.024	(-.073)		
		Total connections	.004	.003
		Causal chain	.021	.000
Mean of all Texts	.080***	(.008)		
		Total connections	.014*	.019**
		Causal chain	.032***	.012*

*p < .05. **p < .01. ***p < .00.

combined (in a manner similar to that used for the importance ratings with total connections and causal chain as factors). This analysis revealed no significant effects. Thus, the causal model did not provide a good fit for recall data for procedural text.

Relationship Between Importance Ratings and Recall

Previous research with narratives has generally shown that text units that are rated more important are recalled better than those rated less important (Brown & Smiley, 1977; Johnson, 1970). To determine whether this relationship occurs for procedural texts, correlations were performed between the importance ratings from Experiment 1 and the recall measure from Experiment 2. These correlations were performed on each individual text for both propositions and idea units and are shown in Table 9. For the Table 7 propositions, the correlation was significant ($p<.05$) for six of the eight texts with a mean correlation of .178 ($p<.001$). For the idea units, the correlation was significant ($p<.05$) for one of the eight texts with a mean correlation of .069 ($p>.05$). The difference between these mean correlation did not quite reach a significant level ($p<.07$).

Table 9

Correlations between Importance Ratings and Mean Recall for Propositions and Idea Units

Text	Propositions	Idea units
Clock	.235*	.073
Form	.354***	.297
Mask	.163*	-.079
Ohmmeter	.108	-.094
Radio	.191*	.163
Recipe	.202*	.298*
Ribbon	.154**	-.061
Spool	-.001	.009
Mean of all texts	.178***	.069

*$p < .05$. **$p < .01$. ***$p < .001$.

GENERAL DISCUSSION

These experiments tested how well the process and causal model predict the importance ratings and the recall of procedural texts. As in previous studies with narratives (Fletcher & Bloom, 1988; O'Brien & Myers, 1987; Trabasso & van den Broek, 1985), the present results with procedural texts showed better predictive power of the causal model for importance ratings. However, neither model performed well with recall. Neither model was as predictive of procedural text as it was in previous research with narratives. In addition, this research shows that for procedural texts, importance ratings and recall are not necessarily related to each other, particularly when idea units are the unit of analysis. In addition, one of the variables, level, has an opposite effect on the two measures.

Ratings

Experiment 1 showed that in general, the causal model accounted for more variance in importance ratings for idea units than the process model did for propositions. The average proportion of variance accounted for by the causal model was .158 compared to .043 for the process model. The causal model accounted for more variance regardless that one less independent predictor was used for this model. The causal model did not account for more variance in the importance ratings than the process model did because of differences in the reliability or variability of ratings for the two types of units (idea unit versus propositions). The variables from the causal model also accounted for more variance in the sentence ratings than the predictors from the process model did (average R^2 = .231 versus .119). This result is important because it shows a pattern of results similar to the previous analyses with an independent set of ratings which is the same for both sets of predictors.

The ANOVA for ratings of the idea units for all texts combined showed causal chain to be a more predictive variable than total connections. This is not surprising since the causal chain tends to include the main cause-and-effect relationships of procedural text.

For the importance ratings, there was no relationship between which texts were significant for the process model and the causal model. Five total R^2 values (of the eight) were significant for each model. Only two texts were significant for both models. Thus, it appears that the predictors from the two models were not related to each other. This is consistent with the findings of Fletcher and Bloom (1988) in which predictors of the two models combined yielded better predictions than either model alone.

Recall

The results from Experiment 2 showed that while there were a few statistically significant R^2 values for the recall data, neither model predicted those data very well. The average proportion of variance accounted for by the causal model was .080 and .037 for the process model. Generally, the R^2 values were smaller for recall than ratings. Trabasso and van den Broek (1985) found that the causal model did not fare as well with recall as with importance ratings, and this was supported in the present study. The process model was expected to do better with recall than with importance ratings; however, this result was not found.

Ratings Versus Recall

Another interesting finding in the present research is the relationship between importance ratings and recall. The results showed that when propositions were the unit of analysis, more of the texts were significant and the average correlation tended to be larger (r = .178) than when idea units were the unit of analysis (r = .069). Based on previous research with narratives that is widely cited in the literature (Brown & Smiley, 1977; Johnson, 1970) the correlations were not as large as expected; however, these results did not come as a complete surprise. Meyer and McConkie (1973), using expository texts, found that with one text, there was virtually no relationship between importance ratings and recall while for another, there was a strong relationship. In the present study, obtaining a strong relationship between ratings and recall would have been surprising given the relationships found between those measures and some of the predictors. For example, the level effect (derived from the process model) found in the ANOVA of the propositions for all texts combined was contradictory for ratings and recall.

As expected for the importance ratings, as level increased (became more subordinate), the ratings showed the propositions were considered less important; however, higher level values were associated with greater recall. This is not consistent with previous research (Kintsch & Keenan, 1973). In the present study, this may have occurred for a combination of reasons. It appears that some of the high level propositions were not recalled because they were assumed, and some of the low level propositions were well recalled because they contained details which were very salient (Birkmire, 1985). While not significant for the idea units, some trends were in opposite directions for ratings and recall. In addition, an inspection of the mean importance ratings and the mean number recalled for idea units suggested why the relationship between ratings and recall was not great. For example, the major actions of the procedure were consistently rated very important; however, many of the major actions were not recalled consistently. Another obvious discrepancy was in the goal statements (the main reason for performing the procedure). These statements tended to be rated low in importance but were well recalled.

Comparison with Previous Research

Generally, the R^2 values in this research were not as large as in the previous research. For example, for importance ratings, Trabasso and Sperry (1985) found the average proportion of variance accounted for by the causal model was .30 and for recall, Miller and Kintsch (1980) found that the average proportion of variance accounted for by five variables related to the process model was .49.

The process model may not be predictive of performance for procedural text for a number of reasons. Procedural text is a different genre of text than the model was developed for, and the model may not be as appropriate with this genre of text. The purpose of procedural text is different than that of narratives or expository text. With procedural text, readers are trying to learn how to perform an action, whereas narratives or expositories are read for other reasons (e.g., for enjoyment or to learn facts [Graesser & Goodman, 1985; Kieras, 1985]). As a result with procedural texts, other processes may be involved (e.g., the construction of a mental model of the procedure that may conceal the microstructure processes proposed by Kintsch).

In addition to the genre differences, additional factors may have contributed to the process model's poor performance in the present studies with procedural text. The present texts tend to have list-like structures. Miller and Kintsch (1980) found that their model failed to predict recall of texts with list-like structures. They stated that "the failure to fit these list-like paragraphs may indicate a basic limitation of the model: It presumes a well-structured text. When it gets a list of sentences that have no text-like structure, it is unable to adapt its strategies and makes plainly wrong predictions." The results in this study confirm their view. One of the reasons that the process model does not perform well with list-like structures is perhaps suggested by these data. Most of the propositions (97%) occur only in one or two cycles of the text analysis and 81% occur only in one cycle, resulting in low variability for the number of cycles variable. However, there is more variability in the levels variable.

The process model may not have performed as well in the present studies as in previous work because of the small number of reinstatements and the lack of inferences. For the texts used by Miller and Kintsch (1980), reinstatements and inferences were the strongest predictor variables. Since these texts had almost no reinstatements or inferences, it is perhaps surprising that the process model performed as well as it did.

In an attempt to increase the number of reinstatements and inferences, this study performed the propositional analyses again of two texts. In these analyses, a proposition was not allowed to be connected with other propositions on the basis of placeholder argument(s). The number of reinstatements increased, but there were no inferences in one of the texts and four inferences in the other. Then, the regression analyses were done with recall as the outcome measure. The proportion of variance accounted for by number of reinstatements was negligible for both these texts. Based on these analyses, it appears that increasing the number of reinstatements in this way did not improve the predictive power of the process model.

The process model would probably achieve better predictive power if the macrostructure level were included in addition to the microstructure. However, the type of macrostructure used in procedural texts and in narrative and expository texts may be different. In any case, in the present study, the process model did not achieve the power that it did in previous studies which used only the microstructure (Kintsch & van Dijk, 1978; Miller & Kintsch, 1980). In addition to differences between the texts used in these experiments and those used by Kintsch and associates, there is an obvious methodological difference that contributed to differences in results. The study used proposition as the unit of analyses, whereas Miller and Kintsch used text as the unit of analyses. This results in a different way of deriving the predictor variables. For example, for the number of cycles variable, Miller and Kintsch counted the total number of cycles needed to construct a coherence graph for a given text, whereas the number of cycles was counted in which a given proposition occurred.

The causal model better predicted the importance ratings than the process model did. This seems logical since the causal model emphasizes the cause-and-effect relationships which are so important in procedural text. As in other studies (Trabasso & van den Broek, 1985), there was a variability in the size of the R^2 values from text to text, as well as in which predictor variable accounted for the most variance. An obvious reason for the smaller amount of variance accounted for by the causal model in the present study than in previous ones is differences in the texts--procedural texts are not as well structured. Also, procedural texts tend to be restricted in the type of relationships that occur in the text. Trabasso and Sperry (1985) discuss six types of relationships that occur in narrative texts. In these experimental texts, only three of these types of relationships occurred. Thus, there was a restriction in the type of relationships, and this may have affected the number of relationships that were possible in the text.

Even though the causal model provided better predictions of the ratings than the process model, neither model accounted for a consistently large proportion of variance, and neither model performed particularly well with the recall data. While there was some variability from text to text, these general conclusions do not appear to be spurious since they are based on a relatively large sample of eight texts. These results are important because it is not clear that either model is sufficient as a representation of procedural text. A model for procedural text needs to take into account not only the text representation, but also the representation of the task. The purpose of reading procedural text of the type used in the present study is to build a cognitive representation of the necessary activities that must be performed. The text representation is different from the cognitive representation of the text's meaning (Dixon, 1987b; Schmalhofer & Glavanov, 1986; van Dijk & Kintsch, 1983). Neither the causal model nor the process model are concerned with the cognitive representation of the task.

Related to this issue is the question of whether task performance would be a better dependent measure for procedural text. The present research used importance ratings and recall to make the research comparable to the earlier work. Of course, task performance is not a possible measure with narratives, but it is certainly appropriate for procedural texts. However, the design of these models does not lend itself to the use of task performance as an outcome.

Because of the limitations of the causal and process models for procedural texts, further work is needed to explore a different type of model which involves variables that are perhaps unique to procedural text, such as the cognitive representation of the task or goals and subgoals of the task. While it is desirable to have one model encompassing all genres of texts, the present results suggest that it is inappropriate to assume that when a model works well for one genre of text, it will work equally well for others. Models that tell something about the comprehension process must be tested for their generality before it is assumed that certain variables are important for all text processing. When limitations of models are discovered, this is important for researchers to know so that existing models can be modified and new, more appropriate models can be developed.

REFERENCES

Birkmire, D. P. (1985). Text processing: The influence of text structure, background knowledge, and purpose. Reading Research Quarterly, 20, 314-326.

Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), Understanding expository text. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bromage, B. K., & Mayer, R. E., (1981). Relationship between what is remembered and creative problem-solving performance in science learning. Journal of Educational Psychology, 73, 451-461.

Brown, A. L., & Smiley. S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognition development. Child Development, 48, 1-8.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Department of the Army. (1976). Mask, chemical-biological: Aircraft ABC-M24 and accessories and mask, chemical-biological: Tank M25/M25A1 and accessories (TM 3-4240-280-10). Washington, DC: Author.

Department of the Army. (1980). Operator's manual radio set AN/PRC-68 (TM 11-5820-882-10). Washington, DC: Author.

Dixon, P. (1987a) The processing of organizational and component step information in written directions. Journal of Memory and Language, 26, 24-35.

Dixon, P. (1987b) The structure of mental plans for following directions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 18-26.

Fletcher, C. R. & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. Journal of Memory and Language, 27, 235-244.

Graesser, A. C. (1978). How to catch a fish: The memory and representation of common procedures. Discourse Processes, 1, 72-89.

Graesser, A. C., & Goodman, S. M. (1985). Implicit knowledge, question answering and the representation of expository text. In B. K. Britton & J. B. Black (Eds.), Understanding expository text. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hay, H. (1975). Learn Magic: The magician's basic tricks explained step-by-step in 20 simple lessons (2nd ed.). New York: Dover Publications.

Herbert, D. (1980). Mr. Wizard's supermarket science. New York: Random House.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. Journal of Verbal Learning and Verbal Behavior, 9, 12-20.

Kieras, D. E. (1985). Thematic processes in the comprehension of technical prose. In B. K. Britton & J. B. Black, J. B. (Eds.), Understanding expository text. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. Journal of Memory and Language, 25, 507-524.

Kieras, D. E., Tibbits, M., & Bovair, S. (1984). How experts and nonexperts operate electronic equipment from instructions (Technical Report No. 14 [UARZ/DP/TR-83/ONR-14]). Tucson: University of Arizona, Psychology Department.

Kintsch, W. (1985). Text processing: A psychological model. In T. A. van Dijk (Ed.), Handbook of discourse analysis, Volume 2. London: Academic Press.

Kintsch, W., & Keenan, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 5, 257-274.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? Journal of Educational Psychology, 65, 109-117.

Meyer, B. J. F., & Rice, G. (1984). The structure of text. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), Handbook of reading research. New York: Longman.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory, 6, 335-354.

Mohammed, M. A. H., & Swales, J. M. (1984). Factors affecting the successful reading of technical instructions. Reading in a Foreign Language, 2, 206-217.

O'Brien, E. J. & Myers, J. L. (1987). The role of causal connections in the retrieval of text. Memory & Cognition, 15, 419-427.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. Memory & Cognition, 14, 64-78.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. Journal of Memory and Language, 25, 279-294.

Spilich, G., Vesonder, G., Chiesi, H., & Voss, J. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 18, 275-290.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), Learning and comprehension of text. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., & Sperry, L. J. (1985). Causal relatedness and importance of story events. Journal of Memory and Language, 24, 595-611.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. Journal of Memory and Language, 24, 612-630.

van den Broek P., & Trabasso, T. (1986). Causal networds versus goal hierarchies in summarizing text. Discourse Processes, 9, 1-15.

van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. Journal of Memory and Language, 27, 1-22.

van Dijk, T. A., & Kintsch, W. (1983). Strategies of Discourse Comprehension. New York: Academic Press.

APPENDIX

PART OF THE CAUSAL ANALYSIS FOR THE RADIO TEXT

PART OF THE CAUSAL ANALYSIS FOR THE RADIO TEXT

(The idea units for the "Radio" text are shown in Table 4 and the diagram of the causal connections is shown in Figure 2).

Units 22 and 33 are temporally coexistent because they are a restatement of each other. Units 23 and 24 are also restatements of each other, and the connection of the handset to the audio connector (23 and 24) enables both the operation of the radio with a standard handset (22) and the turning of the connector fully clockwise (28). Units 25, lining up the two keyways, and 26, pressing down firmly, are temporally coexistent and jointly physically cause the handset to be connected (23 and 24). Units 27 and 29 are temporally coexistent because they are restatements of each other, and they both physically cause the Spkr/Mic to be disconnected (30). Turning the connector fully clockwise (28), physically causes the connector to be locked (27 and 29). In the circumstances, the disconnection of the speaker (30) enables the user to turn on the power (31); in other words, once the speaker is disconnected, the handset is operational, therefore the radio is ready to be used and the power can be turned on. Turning on the power (31), enables one to use the "Push To Talk" switch on the handset (32), and using this switch enables the operation of the radio handset (22 and 33).

Note. Spk/Mic is an abbreviation for speaker microphone.